How Childhood Neighbourhoods Affect the Probability of Matriculating from High School

Arttu Kauhanen

May 5, 2019

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty	Laitos/Institution– Department
Faculty of Social Sciences	Department of Political and Economic Studies

Tekijä/Författare – Author
Arttu Kauhanen

Työn nimi / Arbetets titel – Title
How Childhood Neighbourhoods Affect the Probability of Matriculationg from High School

Oppiaine /Läroämne – Subject
Economics

Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Master's Thesis	May 2019	36 + 16

Tiivistelmä/Referat – Abstract

In my thesis, I estimate the childhood exposure effects of regions in Finland on the probability of completing high school matriculation examination. I estimate the degree to which the differences in high school matriculation rates across regions are driven by the causal effects of places. I study almost 180,000 children who move across regions by exploiting variation in the age of children at the time of the move. I find that neighbourhoods might have a significant childhood exposure effect on girls of low-income families. The outcomes of girls of low-income families change linearly in proportion to the amount of time they spend growing up in a new area at a rate of approximately 6 % per year of exposure. It implies that children who move at birth would pick up 90 % of the difference in permanent residents' outcomes between their origin and destination regions by the age of 16. The results for boys support the critical age model and imply that areas have no childhood exposure effects on boys: the outcomes of boys are unrelated to their age at the time of the move. This implies that the likelihood of boys to complete high school may be unaffected by the families' choices where to live, or boys are affected by the move to a new area at similar magnitude irrespective of the age at the time of the move. The estimation using data of all girls gives a less clear result, which might imply heterogeneity of exposure effects across parents' income levels. The results are robust to alternative specifications and to the overidentification test based on different birth cohorts.

Avainsanat – Nyckelord – Keywords
Intergenerational income mobility, inequality, economic geography, high school matriculation

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty	Laitos/Institution– Department
Valtiotieteellinen tiedekunta	Politiikan ja talouden tutkimuksen laitos

Tekijä/Författare – Author
Arttu Kauhanen

Työn nimi / Arbetets titel – Title
Kuinka lapsuuden asuinalueet vaikuttavat todennäköisyyteen suorittaa ylioppilastutkinto

Oppiaine /Läroämne – Subject
Taloustiede

Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Maisterin tutkielma	Toukokuu 2019	36 + 16

Tiivistelmä/Referat – Abstract

Tässä tutkielmassa arvioin, onko asuinalueilla vaikutusta eri maakunnissa varttuvien lasten todennäköisyyteen suorittaa ylioppilastutkinto ja jos on, kuinka suuria vaikutukset ovat. Työssäni analysoin lähes 180 000:ta maakuntien välillä muuttavaa lasta hyödyntäen lasten iän vaihtelua muuttohetkellä. Tulosten perusteella lapsuuden asuinalueella voi olla merkittävää vaikutusta pienituloisissa perheissä varttuviin tyttöihin. Pienituloisten perheiden tyttöjen lopputulemat muuttuvat lineaarisesti suhteessa uudella alueella vietettyyn aikaan. Todennäköisyys suorittaa ylioppilastutkinto nousee/laskee 6 prosentin vuosivauhtia suhteessa alueiden väliseen eroon tutkintojen suorittamisasteissa. Tällöin ne pienituloisten perheiden tytöt, jotka muuttavat uudelle alueelle heti syntymänsä jälkeen, saavuttaisivat keskimäärin 90 prosenttia alueiden välisestä erosta 16 ikävuoteen mennessä. Poikien lopputulemat taas eivät riipu siitä, missä iässä pojat muuttavat maakuntien välillä. Siten poikien todennäköisyys suorittaa ylioppilastutkinto joko ei riipu perheiden valitsemasta asuinalueesta, tai sitten muutto uudelle alueelle vaikuttaa poikiin yhtä voimakkaasti muuttoiästä riippumatta. Estimointi aineiston kaikilla tytöillä antaa vähemmän selvän tuloksen, mikä voi tarkoittaa vaikutusten olevan riippuvainen vanhempien tulotasosta. Mallin vaihtoehtoiset määrittelytavat tuottavat samankaltaisia tuloksia, ja lasten ikäluokkaan perustuvat yli-identifikaatiotestit tukevat tulosten kausaalista tulkintaa.

Avainsanat – Nyckelord – Keywords
Sukupolvien välinen tuloliikkuvuus, tuloerot, eriarvoisuus, ylioppilastutkinto

Contents

1 Introduction 1

2 Research Question 4

3 Research Methods and Empirical Framework 8

3.1 Important Concepts 8

3.2 Observational Data 10

3.3 Critical Age Effects versus Exposure Effects 13

3.4 Identification Assumption 13

3.5 Other Possible Outcomes 15

4 Empirical Work 16

4.1 Available Data and Sample Definitions 16

4.2 Variable Definitions and Summary Statistics 18

4.3 Estimates of Exposure Effects 22

4.4 Outcome Based Placebo Test 29

5 Conclusions 33

References 35

6 APPENDIX i

List of Tables

1 Frequencies of Movers and Residents 18

2 High School Completion Rates by Family Income Decile (All Movers) 20

3 High School Completion Rates by Family Income Decile (One Time Movers under 16) . . . 21
4 High School Completion Rates by Region . . . 23
5 Regression Results Table . . . 26
6 High School Completion Rates by Family Income Decile and Gender i
7 Frequencies of Residents and Movers by Family Income Decile . . . ii
8 Regression Results: All Children . . . vi
9 Regression Results: Boys . . . vii
10 Regression Results: Girls . . . viii
11 Regression Results: Children of Low Income Families . . . ix
12 Regression Results: Boys of Low Income Families . . . x
13 Regression Results: Girls of Low Income Families . . . xi
14 Frequencies of One-time Movers by the Age at Move . . . xii
15 Regression 7 Results . . . xiii
16 Frequencies of Permanent Residents by Region and Income Decile . xiv
17 High School Matriculation Rates of Permanent Residents by Region and Income Decile . . . xv

List of Figures

1 Estimated Exposure Effects on All Children . . . 25
2 Estimated Exposure Effects on Girls of Low Income Families . . . 27
3 Estimated Exposure Effects on Boys . . . 28
4 Average Outcomes by Birth Cohort in Four Example Regions for Girls of Low Income Families . . . 29
5 Separate Exposure Effects Estimates Based on Cross-Cohort Variation 32
6 Estimated Exposure Effects on Girls . . . iii

7 Estimated Exposure Effects on Boys of Low Income Families iv
8 Estimated Exposure Effects on Children of Low Income Families . . v
9 Matriculation Rates by Birth Cohort and Region xvi

1 Introduction

While income inequality can be seen as a measure of equality of outcome, social mobility and intergenerational income mobility can be seen as a measure of equality of opportunity. Equality of opportunity is affected by many different aspects. Since income mobility varies between countries and areas, one possible factor affecting mobility is childhood environment. An interesting question is whether childhood neighbourhoods matter for outcomes later in life and hence the equality of opportunity? In this study I find significant childhood exposure effects of regions on the probability of girls of low income families to matriculate from high school.

In this work I replicate the article of Chetty and Hendren (2018a) with some modifications using Finnish data. In their article, they estimate the degree to which the differences in inter-generational income mobility across areas are driven by the causal effects of places. In this study I estimate how the probability of completing high school matriculation examination is affected by the causal effect of neighbourhoods where children grow up. The motivation for this kind of a modification is smaller geographic variation in income mobility in Finland compared to the U.S. Small variability in the explanatory variable is likely to hinder precise enough estimation using the Finnish data, whereas high school attendance rates vary more (Ansala, 2018). The use of high school matriculation as an outcome variable is motivated by previous work of Ansala (2018): The study on upper secondary education reveals that children living in a municipality with 10 percentage points higher upper secondary completion rates have a 1.9 percentage points higher likelihood of completing the degree themselves. Children whose parents have lower income respond stronger to their childhood exposure.

It is of interest to conduct this modified replication, because the choice between high school and vocational school correlates with future income and health and has therefore much to do with the heritability of income and social mobility. It is also a choice that divides population into more equal groups instead of affecting only the most disadvantaged children. Hence this choice affects different part of the population than completion of upper secondary education.

The geographic variation in high school matriculation rates could be driven by two different sources: childhood neighbourhoods can have causal effects on the likelihood of completing high school, or there could be systematic differences in the types of families living in each area. I follow Chetty and Hendren (2018a) to compare these two explanations by studying whether children moving to areas with higher (or lower) high school matriculation rates among "permanent residents" are more (or less) likely to matriculate high school themselves. However, since the decision to move is an endogenous choice, simple comparison of the outcomes of moving children confound causal effects of areas with selection effects, i.e. differences in unobservable characteristics of the families cause a bias on the estimates. I address this selection problem by using the age of the children at the time of the move as an instrument variable. With that, I estimate the rate at which the matriculation probabilities of the moving children converge to those of the permanent residents. This identification gives the causal effect of the place, *if* the unobservables of the children do not vary with the age of the child at the time of the move. This identification assumption is strong and can be violated for example if families moving with younger children are more educated. Therefore, I present an outcome based overidentification test to give support to the validity of the estimation. (Chetty and Hendren, 2018a)

Using quasi-experimental research design, Chetty and Hendren (2018a) find that in the US the outcomes of children who move to more affluent neighbourhoods improve linearly in proportion to the time spent in the area. That is, there is no "critical age" effect, rather each year of childhood matters roughly equally. They find that the outcomes of the moving children converge towards the outcomes of the children in the destination area at a rate of about four per cent per year. They also find that place matters mostly through childhood effects, and not because of differences in labour market conditions. Conducting a study using a randomized controlled trial (Moving to Opportunity project) leads to similar results: Children living in families that received a voucher to move to a low poverty area earned substantially more than children whose families did not receive a voucher (Chetty et al., 2016). The effect of the vouchers on adults was statistically insignificant (see for example Kling et al. (2007) or Ludwig et al. (2013)).

In my work, I study children born between 1977 and 1994 who moved across regions between 1996 and 2014. I find that on average, when girls of low income families spend one year in a region with one per cent higher high school matriculation rate increase their likelihood of matriculating themselves from high school by 0.06 per cent. That is, the likelihood of matriculating from high school converge to the likelihood of permanent residents by 6% per year of exposure. I find no significant exposure effect on boys, but the discontinuous jump on matriculation probabilities at the time when outcome is measured is in accordance with the critical age effects. In other words, the effect of the move on boys may be equally large irrespective of the age of the boys at the time of the move. Another possibility is that the areas have no causal effect on boys.

To assess the validity of the identification assumption, I implement a placebo (overidentification) test based on heterogeneity in outcomes across birth cohorts. Since high school matriculation rates vary significantly across birth cohorts, I test whether the outcomes of moving children converge to those of their own birth cohort and are unrelated to the outcomes of preceding and subsequent cohorts. I find statistically significant convergence only towards their own cohort, which would be unlikely to happen due to a selection bias or omitted variables in the model.

The findings of my work imply that childhood neighbourhoods can matter for the economic outcomes of girls of low income families growing up in those areas. These findings are consistent with previous studies on this subject: municipalities matter for upper secondary education (Ansala, 2018) and commuting zones affect the income of the children (Chetty and Hendren, 2018a). However, the heterogenous results across gender is a new finding and needs to be confirmed in subsequent research.

This work is organized as follows. Section 2 describes the research question of interest. Section 3 presents research methods and important concepts for empirical work. Section 4 describes the data and reports the results of the empirical work. Section 5 concludes and section 6 is for appendices. Through my text, I use the concept *better area* for brevity to refer to areas producing higher outcomes and it should have no normative meaning. It takes no stance on how those areas relate in other dimensions which could be more important after all.

2 Research Question

Differences between geographical areas are smaller in Finland than in the United States. In Finland years of education, completion of university degrees and personal income are related to parents' income, especially at the top of the income distribution. This dependence is smaller than in the US, but greater than in other Nordic Countries. There are also regional differences in upward mobility, though these differences are relatively small. For example, between the most mobile and the most immobile areas, children of families at the lowest income decile in Tavastia Proper (Kanta-Häme) region have outcomes at 43–47 percentile point in national income distribution on average, whereas children with similar parental income from Helsinki region or Lapland area have outcomes at around 36–38 percentile points on average. (Suoniemi, 2017)

Despite smaller geographic variation in intergenerational income mobility, the results of Chetty and Hendren (2018a) provoke a question whether childhood environment can have some effect on children's outcomes also in Finland. It is possible that the even quality of Finnish comprehensive schools, free higher-level education and homogenous culture result in places having small or no causal effect on outcomes. However, it is also possible that the areas have causal effects on children's outcomes in Finland, even though the economic significance would be limited compared to the US. This would mean that moving a given child to a different area would change his or her outcome, at least slightly. Since the fraction of children completing high school varies between areas, the question arises whether this variation is causal. If this is the case, it would be of interest to know it. It would also be of interest to obtain information about how large those effects are. Firstly, it would provide information about how much geographical aspects affect the differences in outcomes. If these effects are significant, they should matter for policy decisions. Secondly, it would provide further research topics on what characteristics of the areas cause and/or are related to the differences in the outcomes. This in turn could give a political tool to intervene the heritability of income, health- and educational outcomes, which is often discussed in politics.

Jencks and Mayer (1990) list possible mechanisms why childhood environment could have an effect on children's outcomes. These mechanisms are peer effects, institutional influences and indigenous adult influences. Peer effects mean a social pressure to, for example, finish high school because "everyone else" does. Institutional influences mean influences of adults outside of the community, such as teachers or police. Even if the neighbourhoods per se are irrelevant, better teachers and different treatment of delinquents by police in affluent areas compared to poor neighbourhoods can affect children's life chances. Indigenous adult influences are neighbourhood role models that prove that success is possible if you work hard. Jencks and Mayer (1990) also point out that the definition of neighbourhoods is always geographical rather than social. However, nongeographic communities such as friendship networks can be even more important. If neighbourhoods are heterogenous enough, children are likely to find friends similar to themselves, as most people prefer to do. (Jencks and Mayer, 1990)

Affluent neighbourhoods can also have a negative effect on children's outcomes. People compare themselves to others around them and may judge their success or failure relative to neighbours and friends. The same income would make people feel poorer in a richer area and richer in a poor neighbourhood, and a college dropout could feel less competent in a highly educated area than in an area of high school dropouts. Because of these effects on the opinion of one's own abilities, the most affluent area might not be the best option for children of low income families with less education. Competition for resources such as good grades and jobs can also make an affluent neighbourhood a liability. Chetty and Hendren (2018b) also estimate that the most affluent areas do not produce the best outcomes in the United States. (Jencks and Mayer, 1990)

For estimation purposes, it is useful to classify two possible reasons why children's outcomes can vary between regions. The first is that places have causal effects on economic mobility. In this case, moving children to a different environment would change their outcomes later in their lives. The other possibility is that the differences between the areas are due to the self-selection of different people to live in those areas, which implies that the decision of a family to move from one area to another would have no effect on the outcomes of the children of the family. The purpose of my thesis is to test using Finnish data whether areas indeed have causal effects on children's outcomes, and if so, how large those effects are. (Chetty and Hendren, 2018a)

When making the estimation, I will encounter the same *selection problem* that is encountered when comparing the effects of hospital care. A simple comparison of people who get hospital care to the people who do not may easily yield to a result where patients who got the hospital care have poorer health outcomes than those who refrained. However, it would be misleading to conclude that the effect of the hospital care was indeed negative. In this case, it is easy to see that people who chose not to go to hospital were probably healthier to begin with, and hospitalized people hopefully have better health than they would have in case they had not received the care. The selection bias makes a naïve comparison of these two groups futile, because the *potential outcomes* — a concept that I define precisely in the subsequent chapter — of the groups are different. Since the decision of where to live is also endogenous, the same selection bias is present when comparing the outcomes of the children living in different areas. Unobservables common to households in the same neighbourhoods may be mistakenly attributed to neighbourhood effects (Aaronson, 1998). In the following chapter, I explain how I will tackle this problem.

3 Research Methods and Empirical Framework

3.1 Important Concepts

I first define an important concept for estimation, *potential outcomes*. Potential outcomes are something that we can never observe in reality, but they help to understand the setup. The potential outcome in the area of origin is the outcome the moving children would have obtained, had they lived their entire childhood in the area they moved from. The potential outcome in the area of destination is similarly the outcome the moving children would have obtained, had they lived from the birth until adulthood in the area where they move to. We never get to know neither of them for children who actually move, because neither of the potential outcomes is actually realized. Similarly, we get to know only one of the potential outcomes for children who live in the same area through their entire childhood, the actually realized outcome, but we never know what outcome they would have obtained had they lived in another area. (Chetty and Hendren, 2018a)

I define the *causal effect of the move* as the difference between potential outcomes. Chetty and Hendren (2018a) estimate the potential outcomes using the data of children who live in the same area for their entire childhood, i.e. what the average outcome of children in each area is, given the income of the parents. It is important to condition on parental income, because the income of the parents is an important predictor of the outcomes of the children. Controlling has to be done separately for each area, because place effects can vary across areas at each income level. They also control for birth cohort for each commuting zone, because it is possible that place effects vary over time, and this variation can be different at each income level. I use income percentile *ranks* for the same statistical advantages as Chetty and Hendren (2018a) do.

To elaborate more on this matter, I define the concept of *exposure effect*. The objective of the estimation is to determine how much children's potential outcomes improve on average when they move. Estimating childhood exposure effects answers to this question. I follow Chetty and Hendren (2018a) and define the exposure effect at age m as the impact of spending the year m in an area where the permanent resident's outcomes are one percentile point higher. In an idealised experiment, children who are born in year s are randomly assigned to a new area d starting at age m for the rest of their childhood. The best linear predictor of the outcomes y_i based on the same year s born permanent residents' outcomes $\bar{y}_{pds}$ in parents' income percentile p in the area d would be

$$y_{\mathrm{i}} = \alpha_m + \beta_m \bar{y}_{pds} + \theta_i \tag{1}$$

where θ_i is the error term that is orthogonal to $\bar{y}_{pds}$ due to random assignment and captures the effects of family inputs and other uncontrollable determinants on children's outcomes. Hence β_m gives the *average* effect of spending year m onwards of one's childhood in an area, where permanent residents have one percentile point higher outcomes. It is worth noting that the estimation does *not* give a causal effect of any given area, because these effects are likely to differ between areas. (Chetty and Hendren, 2018a)

The exposure effect at age m is defined as $\gamma_m = \beta_m - \beta_{m+1}$. A positive exposure effect γ_m of any age allows to reject the null hypothesis that areas do not matter. The magnitude of $\sum_{t=0}^{T} \gamma_m$ — the impact of assigning children to a better area from birth — is an estimate of the degree to which the differences in outcomes are due to causal effects vs. selection. $\beta_0 = 0$ would imply that all variation is due to selection effects and $\beta_0 = 1$ would imply that all variation is due to causal effects if place effects are homogenous within birth cohorts. (Chetty and Hendren, 2018a)

3.2 Observational Data

I follow the same identification strategy as Chetty and Hendren (2018a) using Stata for the analysis. They exploit variation in children's ages at the time when their parents move between different areas. Clearly, the parents who move are not representative of the population as a whole, and the decision about whether and where to move is also endogenous. Hence, the error term θ_i in Equation (1) will be correlated with $\bar{y}_{pds}$. For example, parents who move to a richer area may have a latent ability or wealth that affects the outcomes of their children. Therefore, estimating (1) using observational data gives the regression coefficient

$$b_m = \beta_m + \delta_m \tag{2}$$

where $\delta_m = \frac{cov(\theta_i, \bar{y}_{pds})}{var(\bar{y}_{pds})}$ is a selection bias. It measures the extent to which movers' unobservable inputs covary with permanent residents' outcomes in the destination area. However, if the *magnitude* of this selection bias does not vary with children's age at the time of the move, i.e. $\delta_m = \delta$ for all m, the average causal effect of moving for those who move — the treatment on the treated — at time t can be obtained as the difference of the effect of moving at time t and $t+1$, because the selection effect δ cancels out at estimation: $\gamma_m = \beta_m - \beta_{m+1} = b_m - b_{m+1}$. In other words, the families that move to better areas can be different from those who move to worse areas or do not move, as long as the extent of this selection does not vary with the age of the child at the time of the move. (Chetty and

Hendren, 2018a)

Expanding this estimation on all children who move exactly once when they are m years old yields to the following linear regression:

$$y_i = \alpha_{qos} + b_m \Delta_{odps} + \varepsilon_{1i}, \tag{3}$$

where y_i denotes the child's i income rank (or probability of completing high-school matriculation examination), α_{qos} is a fixed effect for the origin area o by parent income decile q and birth cohort s and $\Delta_{odps} = \bar{y}_{pds} - \bar{y}_{pos}$ is the difference in predicted outcomes of permanent residents in the destination versus origin areas given parent income rank p and birth cohort s. Equation (2) can be interpreted as an observational analog of Equation (1).

Expanding regression equation (3) for all ages from 9 to 30 gives:

$$y_i = \alpha_{qosm} + \sum_{m=9}^{30} b_m I(m_i = m)\Delta_{odps} + \sum_{s=1980}^{1987} \kappa_s I(s_i = s)\Delta_{odps} + \varepsilon_{2i}, \tag{4}$$

where α_{qosm} is an origin area o fixed effect by parent income decile q, birth cohort s and age at move m. $I(x_i = x)$ is an indicator function that equals 1 when $x_i = x$ and 0 otherwise. This generalizes (3) by interacting the age at move m with the explanatory variable Δ_{odps}. It also allows the effects of Δ_{odps} to vary across birth cohorts, because the locations of the children are observed from the age 16 for the cohort of 1980, but already from age 8 for the 1988 cohort. This implies greater measurement error in Δ_{odps} for earlier birth cohorts, since their area of origin is undetermined for longer time in their childhood. Even if it is random, a measurement error in an explanatory variable creates a bias in estimates of b_m that is greater for earlier birth cohorts. This bias can be accounted for by coefficients κ_m. The inclusion of cohort interactions means identification from within-cohort variation in ages at move, i.e. comparing children to other children who are born in the same year. (Chetty and Hendren, 2018a)

The problem with (4) is that in the Chetty and Hendren (2018a) study it means more than 200,000 fixed effects (α_{qosm}), which makes it difficult to estimate in small samples and additional controls. Therefore they control parametrically for two main factors of (α_{qosm}) fixed effects: (1) the quality of the origin location and (2) disruption costs of moving. This leads to:

$$\begin{aligned} y_i = &\sum_{s=1980}^{1988} I(s_i = s)(\alpha_s^1 - \alpha_s^2 \bar{y}_{pos}) + \sum_{m=9}^{30} I(m_i = m)(\zeta_m^1 + \zeta_m^2 p_i) \\ &+ \sum_{m=9}^{30} b_m I(m_i = m)\Delta_{odps} + \sum_{s=1980}^{1987} \kappa_s^d I(s_i = s)\Delta_{odps} + \varepsilon_{3i}, \end{aligned} \tag{5}$$

where the first term controls for origin quality by interacting the predicted outcomes of permanent residents in the origin area with birth cohort fixed effects by parent income percentile p_i. The second term controls for the disruption costs of moving. The disruption costs can vary with the age at move and the parent income level. The terms $\{\zeta_m^1\}$ control for children's different ages at the time of the move m. Terms $\{\zeta_m^2 p_i\}$ allow the disruption effects to vary with family income p_i. The third term consists of the exposure effects of interest, and the fourth control for the different measurement errors across birth cohorts, as in equation (3). In their article, specifications (4) and (5) lead to very similar results and therefore it is adequate to control parametrically for the quality of the origin by using permanent residents' outcomes. It implies that the outcomes of the movers can be modelled as a weighted average of the permanent residents' outcomes in the origin and destination using weights that correspond the proportion of childhood spent in two places. These estimates are robust to alternative specifications and sample definitions. (Chetty and Hendren, 2018a)

3.3 Critical Age Effects versus Exposure Effects

There are two possible mechanisms that could explain how areas affect the outcomes of the children: exposure effects that affect in proportion to the exposure time and critical age effects which imply different effects at different ages. In the critical age model, moving to a better area changes children's outcomes with declining probability, but once moved, the exposure time to a more affluent area would not matter for longer term outcomes. These two effects cannot be distinguished in a sample of onetime movers, because of perfect multicollinearity of child's age at move and his exposure time. One possibility to distinguish the two effects is to consider a sample of children who move to a better area, and then move back to where they started. Exposure model predicts that the effect is proportional to the time spent in new area, whereas according to the critical age model the gains depend only on the age at which the child moves to the more affluent area. In their article, Chetty and Hendren (2018a) conclude that their results support for exposure effect model. (Chetty and Hendren, 2018a)

3.4 Identification Assumption

The identification assumption that the selection effects do not vary with the age at the time of the move is strong and can easily be violated. Parents that move to more affluent areas with younger children can be more educated or have otherwise better unobservables than those parents who move with older children. Therefore, Chetty and Hendren (2018a) also provide evidence supporting the validity of the identification in their article using four different approaches. They control for observable fixed family characteristics, time-varying observable characteristics, isolate exogenous moves prompt by aggregate displacement shocks, and implement a set of outcome-based placebo tests. (Chetty and Hendren, 2018a)

The first approach uses siblings within families to estimate how the differences in outcomes are associated with the siblings' age differences interacted with the permanent residents' outcomes in the destination. This approach eliminates confounds that are fixed within families, but do not account for time-varying factors such as changes in family environment at the time of the move that is independent on the living area but affect children in proportion to the exposure time, for example, a divorce or a new job at the time of the move. Hence the second approach controls for changes in income and marital status interacted with the age of the child at the time of the move. The third approach focuses on the moves that are likely to be driven by exogenous aggregate shocks. They identify these shocks by observing large outflows often caused by natural disasters or plant closures. (Chetty and Hendren, 2018a)

All of these first three approaches also rest on assumptions that can be violated. Therefore, the fourth approach relies on a set of overidentification tests that exploit heterogeneity of outcomes across subgroups. The outcomes of different birth cohorts converge to the outcomes of permanent residents of the same birth cohort and are unrelated to the outcomes of other cohorts. This is unlikely to happen due to selection bias. They also exploit variation in the distribution of outcomes across areas. In addition to the mean, also the shape of the outcome distribution of moving children converges to the shape of the outcome distribution in the destination area. Again, it is unlikely that omitted variables could replicate the entire distribution of outcomes in each area in proportion to exposure time. The final placebo test is to use gender specific convergence. If a family moves, for example, to an area that is especially good for boys, their son's outcome is better than their daughter's in proportion to the time spent in that area. It is again unlikely that families are sorting to the areas based on these gender differences. (Chetty and Hendren, 2018a)

Based on these four test, they conclude that "any omitted variable θ_i that generates bias in our estimate of the exposure effect γ must: (1) operate within families in proportion to exposure time (family fixed effect); (2) be orthogonal to changes in parental income and marital status (controls for observables); (3) persist in the presence of moves induced by displacement shocks (displacement shock analysis); and (4) precisely replicate permanent residents' outcomes by birth cohort, quantile, and gender in proportion to exposure time (outcome based blacebo tests)." (Chetty and Hendren, 2018a, p. 33)

3.5 Other Possible Outcomes

In their article, Chetty and Hendren (2018a) focus on income measured outcomes, but also consider other outcomes beyond income: college attendance, marriage, teenage birth and teenage employment. They find that the effect on college attendance is similar to the effect on income (3.7 per cent per year of exposure) and the effect on marriage is slightly smaller (2.5 per cent per year of exposure). The effect on teenage births is specifically large between the ages of 13 and 18. On the other hand, the effect on teenage employment is discontinuous just before the age when the employment is measured. Moving to a higher teenage employment area earlier increases the probability of being employed, but the exposure effect is relatively small compared to the jump. This may suggest that exposure effects consist of different experiences at different points in childhood — such as available summer jobs at certain areas — that may aggregate to produce the linear exposure effects. (Chetty and Hendren, 2018a)

According to Ansala (2018), the areas have an effect on the probability of completing upper secondary education. Children who grow up in a municipality with 10 percentage points higher upper secondary completion rate have 1.9 percentage points higher probability of completing the degree (Ansala, 2018). While the choice not to complete an upper secondary degree is related to social exclusion, the choice between high school and vocational school correlates with future income and health outcomes. High school completion also divides youth to roughly equal sized groups. Therefore, it is of interest to conduct a study about the effect of childhood environment on the likelihood of matriculating from high school. The choice between high school and vocational school is a completely different question than the choice whether to complete an upper secondary education or not, and affects totally different part of the population. The choice between high school and vocational school has major implications on the possibilities later in life and has much to do with the equality of opportunity.

4 Empirical Work

4.1 Available Data and Sample Definitions

I use Finnish Total Statistics on Income Distribution (Tulonjaon kokonaistilasto) from Statistic Finland. The data consist of a register of information from different sources. The outcome variable that I use is a high school matriculation at the age of 20 years. The data include every Finnish person, their postal codes and municipality of residence, personal and household level income data and personal identification codes for each person at each year between 1995 and 2014. However, the link between the parents and the children is missing, and hence I cannot relate the parents' data to the children's data. I can get information about the parents' income through household level income data. However, I have no way to know whether the children move with their parents or move alone for their own purposes, for example to study or to work. Therefore, I have to estimate the effects of moving at ages young enough, when moving is likely to happen with the parents. Another data deficiency is that I cannot relate siblings to each other. Hence I have no means to estimate the effects using within family variation, i.e. comparing the outcomes of the siblings who have different ages at the time of the move.

I follow Chetty and Hendren (2018a) by focusing on variation across wide geographical areas in order to maximize statistical precision in the estimation. They also restrict the analysis to areas with populations above 250,000 to minimize the sampling error on the estimates of permanent residents outcomes. By doing so, I also alleviate the selection issues, because the selection effects are stronger within commuting zones (or regions) compared to between them. Part of this selection problem average out when using larger geographical units for estimation and makes estimation based on a natural experiment more plausible. (Chetty and Hendren, 2018a)

I restrict the data to the individuals born between 1977–1994, of whom there are data available for each of the years 1995–2014. I impose this requirement in order to exclude those children, who have lived abroad during the period of analysis, and to restrict the analysis to those who are alive through the end of 2014. I further restrict the data to individuals who either live in one region the whole period of 1995–2014 (residents), or move between regions exactly once (one-time movers). The individuals who move between regions more than once are excluded from the analysis. In the working sample there are 600,384 residents, 179,504 one-time movers and 218,712 children who move more than once (Table 1).

Table 1: Frequencies of Movers and Residents

	count
Residents	600384
One-time Movers	179504
Several Moves	218712
Total	998600

4.2 Variable Definitions and Summary Statistics

In this section, I define the variables I use in the analysis. I convert all the monetary variables in 2014 euros to adjust for inflation using consumer price index (CPI). My key variables that I use for the estimation are:

Family Income: For the measure of family income, I calculate the total household equivalence income during the years 1995-1999. Because I measure family income for a fixed set of years, the age of the child when family income is measured varies across birth cohorts. As in Chetty and Hendren (2018a) article, I account for this variation by controlling for the child's birth cohort in the analysis. Household equivalent income is measured by OECD-modified scale. When calculating the size of the household, this scale assigns a value of 1 to the household head, of 0.5 to each additional adult member (over 13 years old) and of 0.3 to each child. The size of the household is measured at the last day of the year. The equivalence income is defined as the total net income after transfers and taxes of the household divided by the modified size of the household.

Family Income Deciles: Based on this income measure, I rank all the households at a nationwide level and classify them into ten equal sized income deciles.

High School Completion/Matriculation: By high school completion I mean that a given individual has completed high school matriculation examination by the age of 20 years, and this is the highest degree completed at the age of 20. I use this definition, because the data gives only the highest completed degree for each given year. It is unlikely that a given person would have any higher degree at the age of 20. The people who would matriculate at older ages are considered to have no high school degree throughout the analysis.

As we can see from Table 2, the average high school matriculation rates vary significantly between family income deciles. Only 25 per cent of the children of the families in the lowest income decile have matriculated from the high school, compared to just over 72 per cent of the children of the families in the highest income decile. Matriculation rate of all children is 39.7 per cent. Matriculation rates of the one-time movers and rates of the children who move several times between regions are significantly higher. However, Table 3 reveals that calculating matriculation rates of the one-time movers who move before age 16 yields to very similar rates to the rates of the residents at all income levels. This confirms the intuition that moving at older ages is dominated by people who move between region for academic studies or higher level job opportunities. Hence the selection effects are stronger at older ages, especially for movers who are around 20 years old.

Age: The age of an individual is defined as the age in full years on the last day of the calendar year.

Location: In each year, all individuals are assigned municipality codes according to which municipality they live in on the last day of the year. Based on this information, I classify all the individuals into 18 regions (excluding Ahvenanmaa region) using Statistics Finland 2010 division.

As we can see from Table 4, the lowest high school completion rate is in Kainuu

Table 2: High School Completion Rates by Family Income Decile (All Movers)

	Residents	One-time movers	Several moves
1	.2515943	.4251585	.4161226
2	.2541326	.4352197	.4216138
3	.2880225	.4988877	.4760282
4	.3289033	.5531565	.5354442
5	.3713399	.6049371	.5788263
6	.4116496	.6561836	.6246567
7	.4663662	.6920052	.6764422
8	.527499	.7415062	.7241359
9	.6120913	.7935235	.7793249
10	.7222525	.8462506	.8348736
Total	.3970909	.6034462	.5744084
Observations	600384	179504	218712

Table 3: High School Completion Rates by Family Income Decile (One Time Movers under 16)

	Residents		Movers at age under 16	
	mean	Standard Deviation	mean	Standard Deviation
1	.2515943	.4339335	.2405221	.4275334
2	.2541326	.4353757	.230162	.4210239
3	.2880225	.4528446	.2812131	.4496798
4	.3289033	.4698179	.3290816	.4699996
5	.3713399	.4831665	.4026299	.4905811
6	.4116496	.492136	.4631353	.4988177
7	.4663662	.4988716	.4971182	.5001719
8	.527499	.4992481	.5677817	.4956025
9	.6120913	.4872795	.6278119	.4836355
10	.7222525	.4478943	.7108141	.4536601
Total	.3970909	.4892955	.3867627	.4870239
Observations	600384		15834	

region. Just 25.7 per cent of the children who have grown their entire childhood in Kainuu region have matriculated from high school. The highest rate, 52.6 per cent is in Uusimaa region. The nationwide high school completion rate is 39.7 per cent. I conclude that there is significant variation in high school matriculation rates across regions. The high school completion rate of all the families living in Kainuu region is close to the nationwide completion rate of children whose family income is in the second income decile. The significant regional variation further supports the use of high school matriculation rate as the outcome variable in the analysis.

4.3 Estimates of Exposure Effects

In this section I present the parametric estimates of the exposure effects $\{\gamma_m\}$ based on the following equation:

$$\begin{aligned} y_i = &\sum_{s=1977}^{1994} I(s_i = s)(\alpha_s^1 - \alpha_s^2 \bar{y}_{pos}) + \sum_{m=2}^{30} I(m_i = m)(\zeta_m^1 + \zeta_m^2 p_i) \\ &+ \sum_{m=2}^{30} b_m I(m_i = m)\Delta_{odps} + \sum_{s=1977}^{1994} \kappa_s^d I(s_i = s)\Delta_{odps} + \varepsilon_{3i}, \end{aligned} \tag{6}$$

where $\bar{y}_{pos}$ is the average high school completion rate in the origin region and $\Delta_{odps} = \bar{y}_{pds} - \bar{y}_{pos}$ is the difference in the predicted high school completion rates of permanent residents in the destination versus origin areas given the parent income decile p and the birth cohort s. This is equivalent to the equation (5) with a different outcome variable, birth cohorts and possible ages at the time of the move. I restrict the analysis to the children whose age at the time of the move is less than 16 years. After that point the decision to move is more likely to be endogenous. In other words, children who move between the ages 16–19 are more

Table 4: High School Completion Rates by Region

	mean	Standard Deviation
Etelä-Karjala	.28387	.4508918
Etelä-Pohjanmaa	.2951506	.4561202
Etelä-Savo	.2919614	.4546819
Kainuu	.2574053	.4372328
Kanta-Häme	.2824704	.4502153
Keski-Pohjanmaa	.3050801	.4604726
Keski-Suomi	.335183	.4720624
Kymenlaakso	.2540181	.4353198
Lappi	.3035318	.4597944
Pirkanmaa	.3960805	.4890864
Pohjanmaa	.3618778	.4805543
Pohjois-Karjala	.3470814	.4760546
Pohjois-Pohjanmaa	.3647957	.4813775
Pohjois-Savo	.3138178	.4640522
Päijät-Hme	.2891069	.4533593
Satakunta	.3143345	.4642591
Uusimaa	.5259382	.4993281
Varsinais-Suomi	.4213564	.493781
Total	.3970909	.4892955
Observations	600384	

likely to move because of going to high school. A hint of possible increase in the selection motive can be seen from table 14. The frequencies of moving children are significantly higher between the ages 16–19. This kind of selection would induce a higher upward bias δ_m on the estimates of b_m from Equation (2). I also present results after the age of 20, because a move at an age higher than 20 years should have no effect on the probability of completing high school by the age of 20 years. Therefore, the level of these estimates is estimating the selection effects.

Figure 1 plots the coefficients $\{b_m\}$ from equation (5). The regression coefficient of $b_2 = 0.766$ estimated from Equation (6) implies that children who move at the age of two years old to an area which has 1 percentile point higher high school matriculation rate $\bar{y}_{pds}$ in the same income decile p and the same birth cohort s, are in average 0.766 percentile points more likely to go to high school than children who do not move (Chetty and Hendren, 2018a). The estimate b_2 consists of the exposure effect β_2 and the selection bias δ_2 equivalent to Equation (2).

Regressing the coefficients b_m on the age at move m for $m \leq 15$ gives the average annual exposure effect estimate of $\gamma = 0.0120$ (s.e. $= 0.0095$), which is statistically insignificant (Table 5). The pattern of the coefficients b_m suggests that the identification assumption might not hold, meaning the *selection effects* $\{\delta_m\}$ from Equation (2) may vary with the age at move m. Parents might be more eager to move when children are at certain ages, namely when children have to change to junior high school for the last three years of comprehensive school (ages 13–15). Parents who do this kind of timing might be different from other parents, and their children might be more likely to go to high school to begin with. This kind of sorting would result in a higher selection bias for age around 13. Another possibility is that the variation is random due to relatively small data sample, which leads to imprecise estimates of the coefficients b_m. Third possibility is that the selection

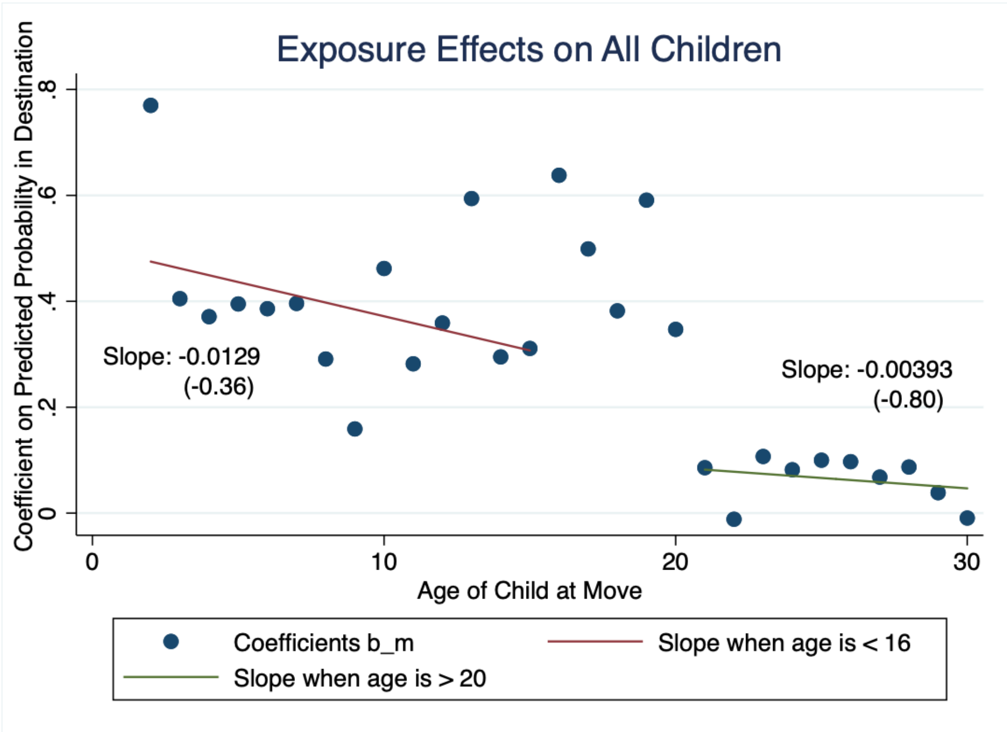

Figure 1: Estimated Exposure Effects on All Children

or the exposure effects are heterogenous across population subgroups. Fourth possibility is that some ages may be more important than others, for example the period of junior high school (ages 13–15) may be more important than earlier years implying critical age effects. This would result in nonlinearities in the model. I have no means to test which of these explanations is true, but based on subsequent analysis I think that heterogenous selection and exposure effects across population subgroups are the most plausible explanation.

Table 5: Regression Results Table

Model:	All	Low inc.	Boys	Girls	Low inc. Boys	Low inc. Girls
Exposure	-0.0129	-0.0288	0.000386	-0.0235	0.0113	-0.0602**
Effect	(-1.36)	(-2.06)	(0.04)	(-1.77)	(0.72)	(-3.21)
Constant	0.501***	0.682***	0.410**	0.516**	0.532**	0.610**
	(5.62)	(5.18)	(4.10)	(4.13)	(3.63)	(3.46)
N	14	14	14	14	14	14

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table (5) summarizes the regression results for the different subpopulations: all observations, children of low income families (deciles 1–3), boys, girls, boys of low income families and girls of low income families. We can see from the table that the result for girls of low income families is statisticlly significant at one per cent confidence level and gives an average annual exposure effect of -0.0602. It means that if the identification assumption holds, the probability of matriculating from high school of girls of low income families who move converge to the probability of permanent residents of the destination area at a rate of 6.02 % per year of exposure

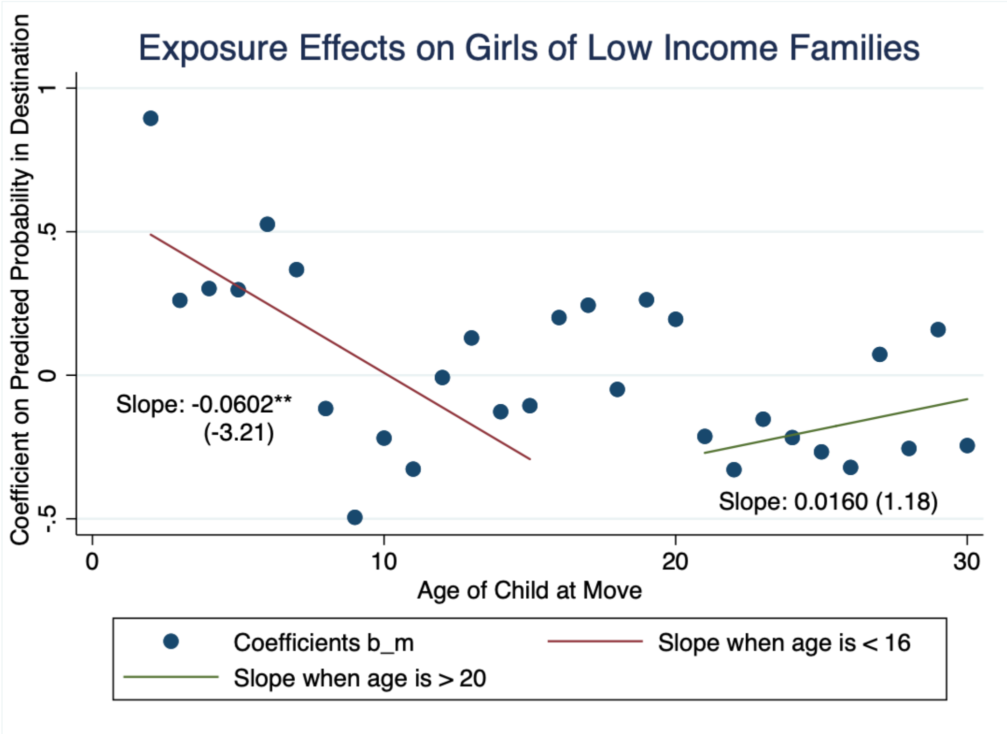

Figure 2: Estimated Exposure Effects on Girls of Low Income Families

until the age of 16 (Chetty and Hendren, 2018a). A visual inspection of Figure 2 reveals a regular pattern of the coefficients $\{b_m\}$ around the regression line, which is in accordance with the identification assumption.

(Figure 3) reveals an interesting result. The average annual exposure effect estimate for boys is 0.000386, which is statistically insignificant. This time the coefficients $\{b_m\}$ exhibit somewhat more regular pattern around the regression line, which is in accordance with the identification assumption that the *selection effects* $\{\delta_m\}$ are orthogonal to the age of the child at the time of the move m. If the identification assumption holds, the result means that boys are unaffected by the average high

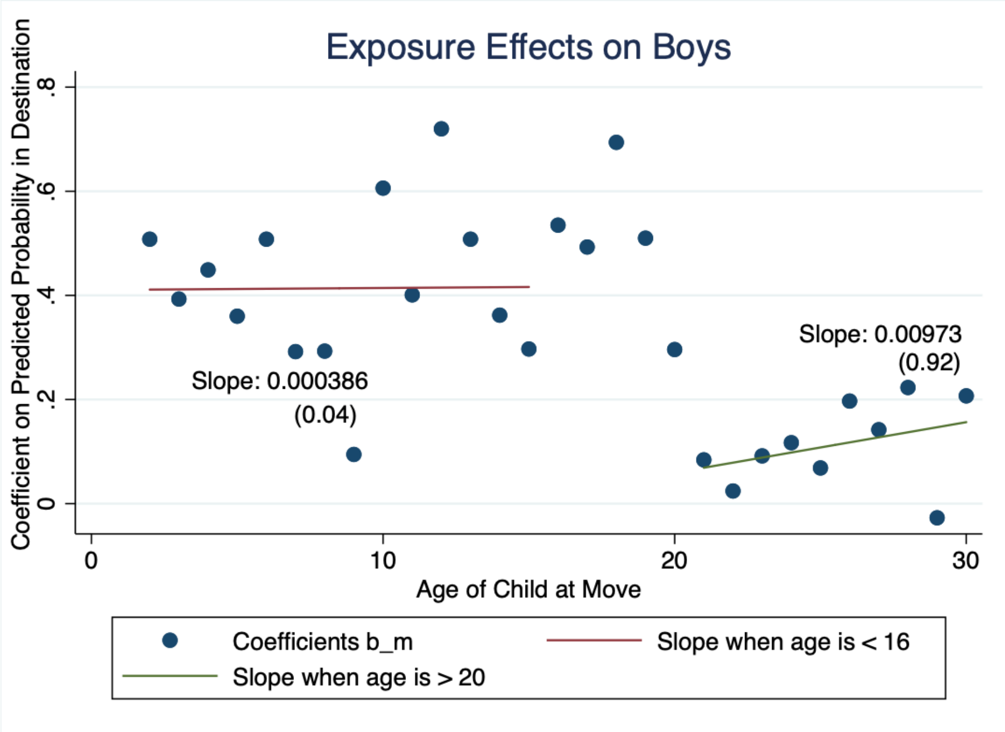

Figure 3: Estimated Exposure Effects on Boys

school completion rates of the region, or the results support the critical age model due to the discontinuous jump when the outcome is measured at 20 years. On the contrary, regions of residence seem to have childhood exposure effects for girls of low income families. The figures for the girls (6), boys of low income families (7) and all children of low income families(8) are presented in the Appendix. The regular pattern of the coefficients $\{b_m\}$ in Figure (6) with almost significant slope and significant slope for girls of low income families gives a hint that girls in general may be affected by the region they live in. The reason for the insignificant result may lie on the low power of the test due to a small data sample.

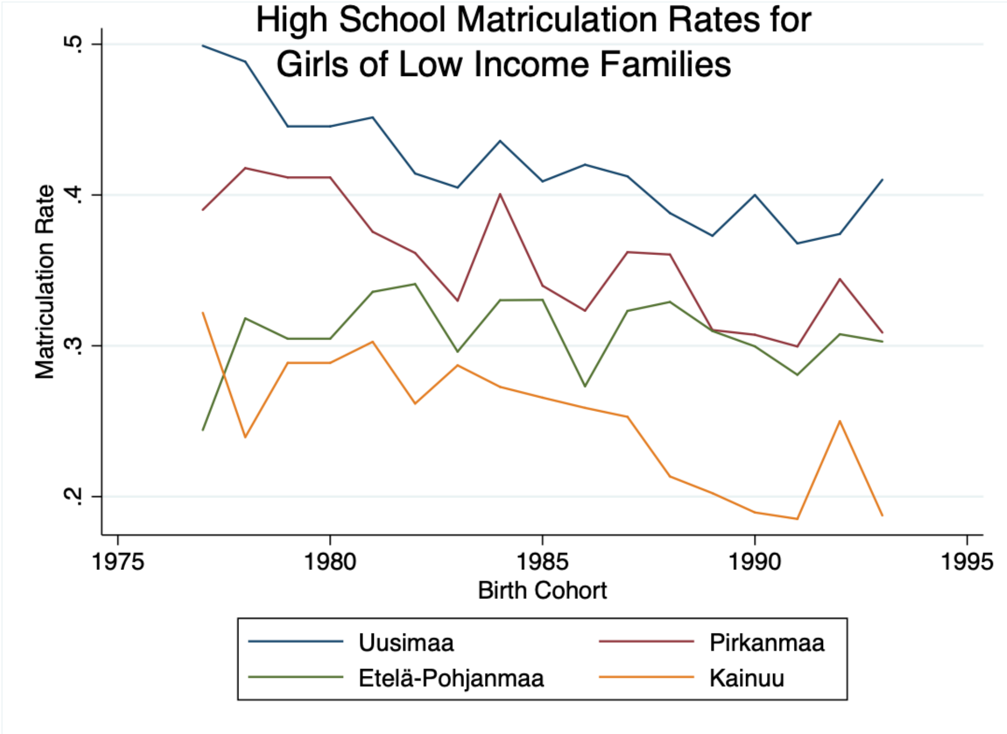

Figure 4: Average Outcomes by Birth Cohort in Four Example Regions for Girls of Low Income Families

4.4 Outcome Based Placebo Test

In order to assess whether the identification assumption holds and to test for bias due to unobservable factors, I implement one placebo test that exploits variation across different birth cohorts. First I present a figure that plots how the outcomes have evolved across birth cohorts in different regions for girls of low income families. As we can see from Figure (4), the average outcomes have significant variation over time in the four example regions presented in the figure. The matriculation rates have abrupt changes in several years. Figure (9) in the Appendix presents how average matriculation rates have evolved across birth cohorts by region. Similar pattern can be found in all the regions.

The test is based on the idea that a neighbourhoods quality for a child's own birth cohort matters, rather than the neighbourhood's quality for older or younger cohorts. If the exposure effect estimates indeed reflect causal relations, a child's outcome should converge towards the outcomes of his or her own cohort in the destination region. In contrast, it is unlikely that an omitted variable bias would generate such a cohort specific convergence. (Chetty and Hendren, 2018a)

In order to conduct the outcome based placebo test, I follow Chetty and Hendren (2018a) and parameterize the exposure and the selection effects shown in Figure(2) by replacing the non-parametric $\sum_{m=2}^{30} b_m I(m_i = m)\Delta_{odps}$ term in Equation (6) with two separate linear terms, one below the age of 16 and another for ages over 19:

$$\begin{aligned} y_i = \sum_{s=1977}^{1994} I(s_i = s)(\alpha_s^1 - \alpha_s^2 \bar{y}_{pos}) + \sum_{m=2}^{30} I(m_i = m)(\zeta_m^1 + \zeta_m^2 p_i) \\ + \sum_{s=1977}^{1994} \kappa_s^d I(s_i = s)\Delta_{odps} \\ + I(m_i \leq 15)(b_0 + (20 - m_i)\gamma)\Delta_{odps} + I(m_i \geq 20)(\gamma + (20 - m_i)\gamma')\Delta_{odps} + \varepsilon_{3i} \end{aligned} \tag{7}$$

This specification yields to similar results to the non-parametric estimates (6). The results are shown in Table (15) in the Appendix. Based on this regression, I estimate the exposure effects by replacing the difference in permanent resident's outcomes Δ_{odps} in child's own birth cohort $s(i)$ for the difference $\Delta_{odp,s(i)+t}$ in other birth cohorts $s(i) + t$. Figure 5 plots the exposure effect estimates $\{\gamma_t'\}$ for t ranging from -4 to 4. The coefficients γ_t' are similar for t from -2 to 0, but the coefficients for $t \neq 0$ are statistically insignificant and get smaller for other t, even though the outcomes of the adjacent cohorts must be somewhat autocorrelated. Despite of this autocorrelation, the variation over time allows to conclude that the outcomes of moving girls of low income families converge to the outcomes of their own birth cohort and are unrelated to the outcomes of other cohorts. This supports the view that the change in moving children's outcomes is driven by causal exposure effects to different neighbourhoods. It is unlikely that families' unobservable characteristics would covary with children's outcomes on the birth cohort level to create such a precice birth cohort specific convergence (Chetty and Hendren, 2018a).

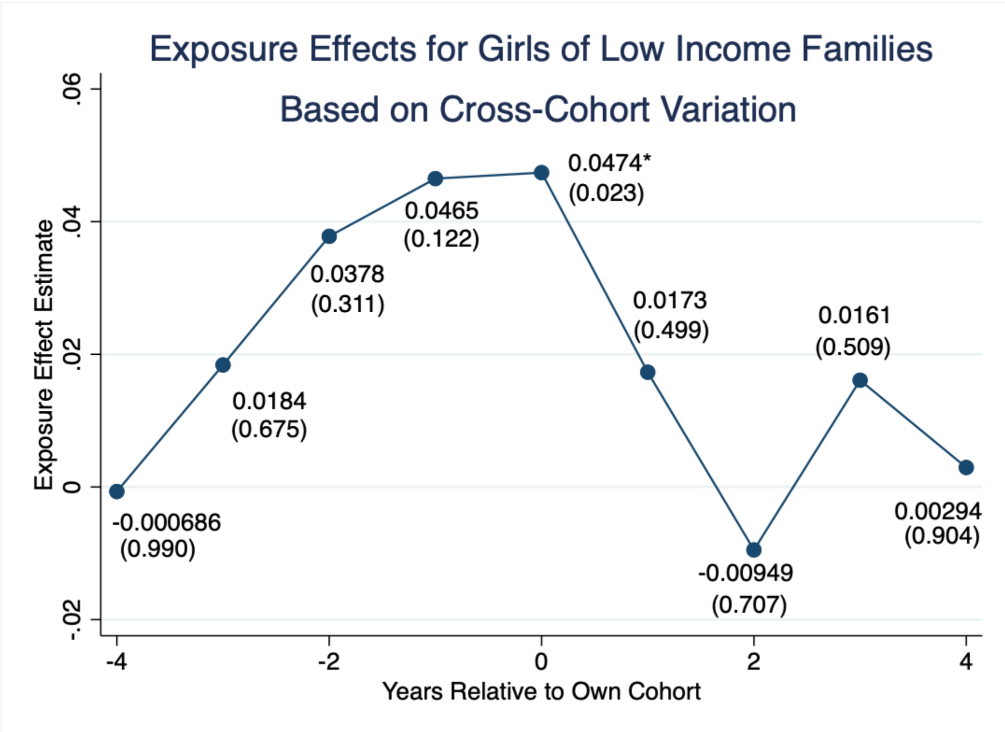

Figure 5: Separate Exposure Effects Estimates Based on Cross-Cohort Variation
Note: Exposure effect of 0.0474 corresponds to the exposure effect estimate in Table 5. The slope of -0.0602 in Figure 2 would correspond to an exposure effect of 0.0602, since a negative slope implies a positive exposure effect.

5 Conclusions

In this work I have found heterogenous childhood exposure effects across gender. The results for boys can have two different explanations: either the results support the critical age model, or boys are unaffected by childhood neighbourhoods. At any case, areas seem to have no exposure effects on boys. Girls of low income families respond strongly to differences in childhood neighbourhoods. The results for girls of low income families support for childhood exposure effects model: the outcomes of moving girls of low income families converge to those of permanent residents in the region to which they move at a rate of approximately 6% per year of childhood exposure. It implies that children who move at birth would pick up 90% of the difference in permanent residents' outcomes between their origin and destination regions by the age of 16.

I have corrected for the problem of selection bias by exploiting variation in children's ages when they move between regions. The results are robust to alternative specifications and to the overidentification test based on variation in outcomes of different birth cohorts. The outcomes of moving girls of low income families converge to the outcomes of permanent residents in the destination in their own birth cohort, and are unrelated to the outcomes of other birth cohorts. It is unlikely that unobservable characteristics of moving families would generate such a cohort specific convergence, suggesting that neighbourhoods may indeed have causal effects on the probability of girls of low income families to matriculate from high school.

The differences in permanent residents' high school completion rates are predictive of regions' causal effects *on average* for girls of low income families. It does not estimate the causal effect of any particular region, as the high school completion rates of permanent residents reflect a different combination of selection and causal effects in each region.

The results are in accordance with previous studies (see for example Ansala (2018), Chetty and Hendren (2018a), Chetty and Hendren (2018b), Chetty et al. (2016), Suoniemi (2017)) in the sense that neighbourhoods seem to matter for the outcomes of the children growing up in those areas. However, different results for boys and girls is a new finding and needs to be confirmed in subsequent studies.

References

Daniel Aaronson. Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources*, pages 915–946, 1998.

Laura Ansala. Essays on the economics of migration and geography; esseitä muuttoliikkeen ja maantieteen taloustieteestä, 2018. ISSN 1799-4942 (electronic); 1799-4934 (printed); 1799-4934 (ISSN-L). URL http://urn.fi/URN:ISBN:978-952-60-8208-0.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162, 2018a.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228, 2018b.

Raj Chetty, Nathaniel Hendren, and Lawrence F Katz. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902, 2016.

Christopher Jencks and Susan E Mayer. The social consequences of growing up in a poor neighborhood. *Inner-city poverty in the United States*, 111:186, 1990.

Jeffrey R Kling, Jeffrey B Liebman, and Lawrence F Katz. Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119, 2007.

Jens Ludwig, Greg J Duncan, Lisa A Gennetian, Lawrence F Katz, Ronald C Kessler, Jeffrey R Kling, and Lisa Sanbonmatsu. Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American Economic Review*, 103(3):226–31, 2013.

Ilpo Suoniemi. Intergenerational mobility and equal opportunity, evidence from Finland. *Labour Institute for Economic Research*, Working Papers(312), 2017.

6 APPENDIX

Table 6: High School Completion Rates by Family Income Decile and Gender

	Residents		All Movers	
	Boys	Girls	Boys	Girls
1	.1837854	.3540361	.3228397	.465865
2	.1879801	.3540677	.3281577	.4817423
3	.2132244	.4096646	.3762041	.5450753
4	.2474552	.4672716	.4266309	.6046081
5	.2842933	.519523	.4799862	.6486508
6	.3208127	.5647718	.529534	.7015473
7	.374115	.6200022	.5878772	.7383058
8	.4355195	.6751701	.6483447	.7830178
9	.5308013	.7374042	.7197008	.8308373
10	.6594695	.8072335	.801329	.8712977
Total	.3198198	.5264929	.4915332	.6369548
Observations	330211	178293	270173	219923

Table 7: Frequencies of Residents and Movers by Family Income Decile

	Residents	One-time movers	Several moves	Total
1	51591	16408	23892	91891
2	68783	20392	27575	116750
3	76001	22476	28471	126948
4	75469	21860	26845	124174
5	71891	21146	25220	118257
6	67453	20223	23304	110980
7	60133	18312	20732	99177
8	51020	15629	16983	83632
9	41319	12723	14220	68262
10	36724	10335	11470	58529
Total	600384	179504	218712	998600
N	998600			

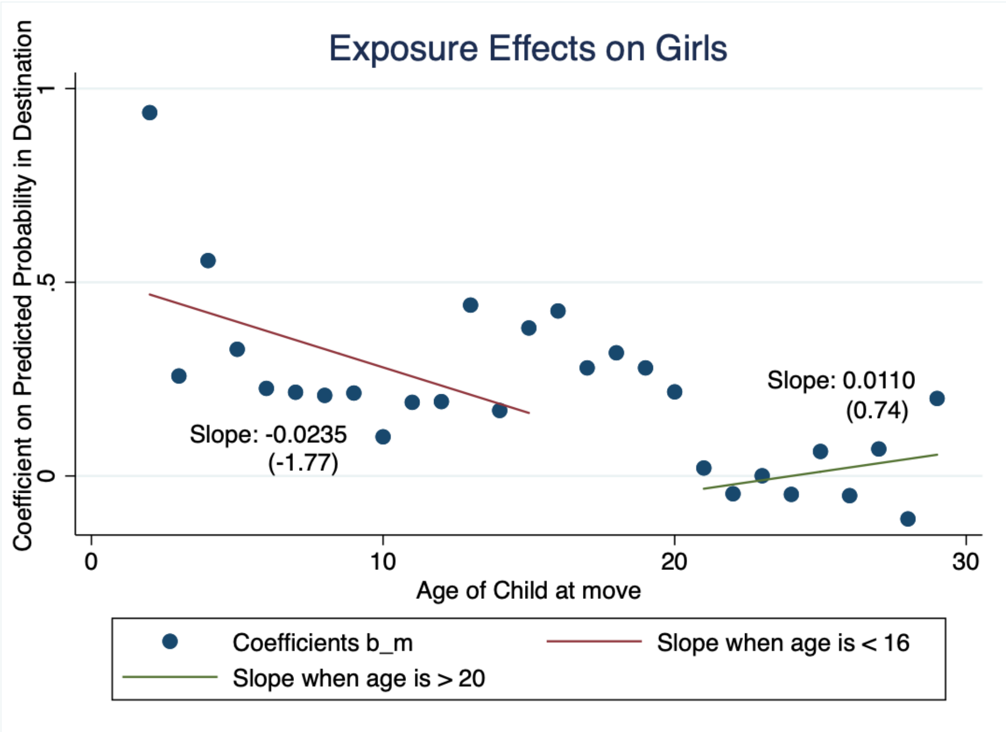

Figure 6: Estimated Exposure Effects on Girls

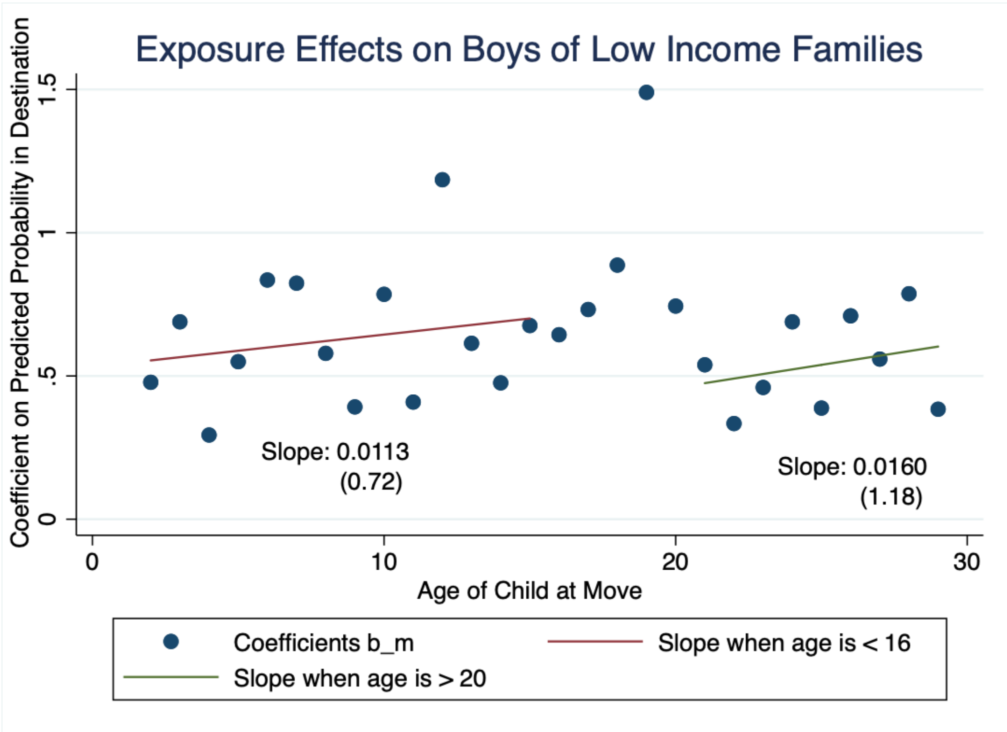

Figure 7: Estimated Exposure Effects on Boys of Low Income Families

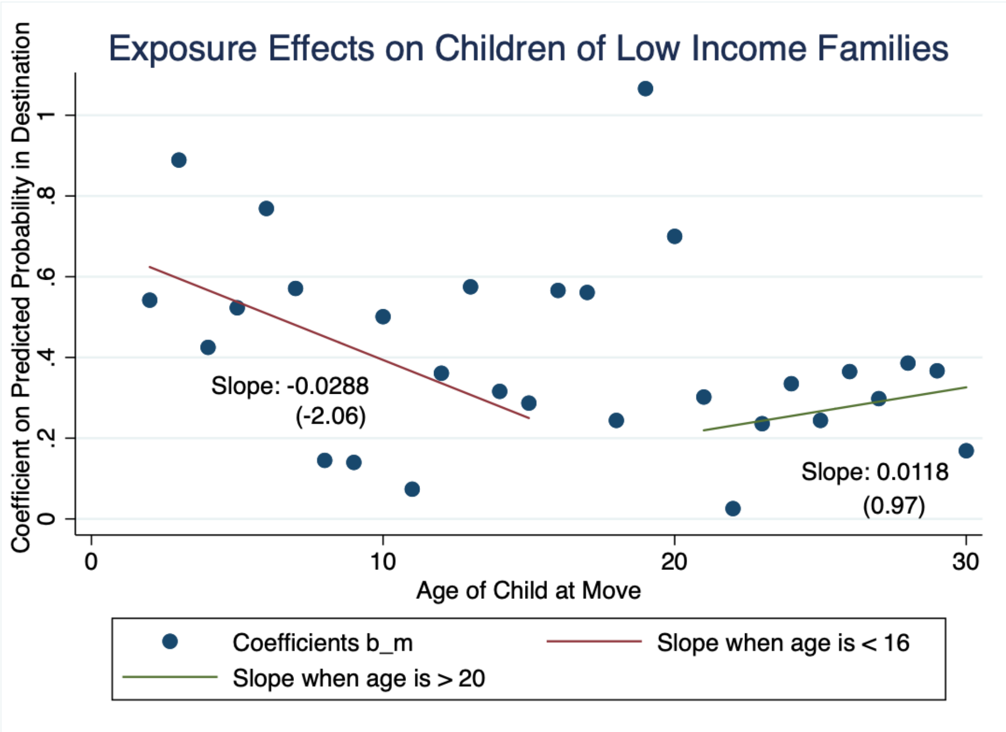

Figure 8: Estimated Exposure Effects on Children of Low Income Families

Table 8: Regression Results: All Children

	Estimated Coefficients b_m	
Age at Move	b_m	p-value
2	0.770***	(0.000)
3	0.405*	(0.016)
4	0.371*	(0.015)
5	0.395**	(0.003)
6	0.386**	(0.003)
7	0.396**	(0.001)
8	0.291*	(0.027)
9	0.159	(0.184)
10	0.462***	(0.000)
11	0.282*	(0.017)
12	0.359**	(0.003)
13	0.594***	(0.000)
14	0.295*	(0.018)
15	0.311*	(0.011)
16	0.638***	(0.000)
17	0.499***	(0.000)
18	0.382***	(0.000)
19	0.591***	(0.000)
20	0.347***	(0.000)
21	0.0858	(0.095)
22	-0.0115	(0.829)
23	0.107*	(0.046)
24	0.0819	(0.122)
25	0.1000	(0.060)
26	0.0972	(0.070)
27	0.0681	(0.217)
28	0.0871	(0.125)
29	0.0386	(0.518)
30	-0.00926	(0.882)
Observations	368130	
F	164.9	
df_m	370	
df_r	368129	

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Regression Results: Boys

	Estimated Coefficients b_m	
Age at Move	b_m	p-value
2	0.508	(0.079)
3	0.393	(0.067)
4	0.449*	(0.023)
5	0.360*	(0.039)
6	0.508**	(0.002)
7	0.292	(0.088)
8	0.293	(0.104)
9	0.0945	(0.573)
10	0.606***	(0.001)
11	0.401*	(0.020)
12	0.720***	(0.000)
13	0.508**	(0.003)
14	0.362*	(0.049)
15	0.297	(0.103)
16	0.535***	(0.000)
17	0.493**	(0.003)
18	0.694***	(0.000)
19	0.510***	(0.000)
20	0.296***	(0.000)
21	0.0841	(0.327)
22	0.0243	(0.789)
23	0.0917	(0.323)
24	0.117	(0.211)
25	0.0685	(0.469)
26	0.197*	(0.040)
27	0.142	(0.156)
28	0.223*	(0.034)
29	-0.0272	(0.810)
30	0.207	(0.073)
Observations	80390	
F	47.33	
df_m	370	
df_r	80389	

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Regression Results: Girls

	Estimated Coefficients b_m	
Age at Move	b_m	p-value
2	0.938***	(0.000)
3	0.258	(0.206)
4	0.556**	(0.002)
5	0.327	(0.069)
6	0.226	(0.218)
7	0.216	(0.185)
8	0.208	(0.242)
9	0.214	(0.180)
10	0.101	(0.547)
11	0.190	(0.253)
12	0.192	(0.227)
13	0.441**	(0.006)
14	0.169	(0.351)
15	0.382*	(0.024)
16	0.426***	(0.000)
17	0.279*	(0.022)
18	0.318***	(0.000)
19	0.279***	(0.000)
20	0.217***	(0.000)
21	0.0205	(0.766)
22	-0.0460	(0.526)
23	0.000389	(0.996)
24	-0.0476	(0.517)
25	0.0632	(0.405)
26	-0.0509	(0.528)
27	0.0696	(0.427)
28	-0.111	(0.256)
29	0.200	(0.061)
Observations	92018	
F	48.74	
df_m	359	
df_r	92017	

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Regression Results: Children of Low Income Families

	Estimated Coefficients b_m	
Age at Move	b_m	p-value
2	0.542	(0.156)
3	0.889**	(0.005)
4	0.425	(0.130)
5	0.523*	(0.041)
6	0.769**	(0.001)
7	0.571**	(0.008)
8	0.145	(0.539)
9	0.140	(0.512)
10	0.501*	(0.023)
11	0.0737	(0.727)
12	0.361	(0.084)
13	0.575**	(0.004)
14	0.316	(0.143)
15	0.287	(0.160)
16	0.566**	(0.002)
17	0.561***	(0.001)
18	0.244	(0.086)
19	1.066***	(0.000)
20	0.700***	(0.000)
21	0.302*	(0.022)
22	0.0256	(0.849)
23	0.236	(0.081)
24	0.335*	(0.013)
25	0.244	(0.074)
26	0.365**	(0.008)
27	0.298*	(0.035)
28	0.386**	(0.007)
29	0.367*	(0.013)
30	0.169	(0.263)
Observations	127044	
F	62.29	
df_m	167	
df_r	127043	

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Regression Results: Boys of Low Income Families

	Estimated Coefficients b_m	
Age at Move	b_m	p-value
2	0.478	(0.435)
3	0.689	(0.160)
4	0.294	(0.465)
5	0.550	(0.151)
6	0.835^{*}	(0.026)
7	0.824^{*}	(0.023)
8	0.579	(0.154)
9	0.392	(0.244)
10	0.785^{*}	(0.038)
11	0.409	(0.256)
12	1.185^{***}	(0.000)
13	0.614	(0.091)
14	0.476	(0.174)
15	0.676	(0.074)
16	0.644	(0.059)
17	0.732^{*}	(0.039)
18	0.887^{**}	(0.003)
19	1.490^{***}	(0.000)
20	0.744^{**}	(0.005)
21	0.539^{*}	(0.049)
22	0.334	(0.232)
23	0.460	(0.105)
24	0.689^{*}	(0.016)
25	0.388	(0.179)
26	0.710^{*}	(0.015)
27	0.559	(0.061)
28	0.787^{**}	(0.009)
29	0.384	(0.225)
Observations	24792	
F	17.79	
df_m	163	
df_r	24791	

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Regression Results: Girls of Low Income Families

	Estimated Coefficients b_m	
Age at Move	b_m	p-value
2	0.895*	(0.049)
3	0.261	(0.472)
4	0.302	(0.373)
5	0.298	(0.368)
6	0.526	(0.087)
7	0.368	(0.198)
8	-0.116	(0.703)
9	-0.495	(0.080)
10	-0.219	(0.461)
11	-0.327	(0.262)
12	-0.00786	(0.978)
13	0.130	(0.626)
14	-0.127	(0.673)
15	-0.106	(0.732)
16	0.201	(0.383)
17	0.244	(0.252)
18	-0.0493	(0.775)
19	0.263	(0.089)
20	0.195	(0.197)
21	-0.213	(0.204)
22	-0.329	(0.061)
23	-0.153	(0.392)
24	-0.217	(0.238)
25	-0.267	(0.163)
26	-0.321	(0.113)
27	0.0726	(0.741)
28	-0.255	(0.283)
29	0.159	(0.519)
30	-0.245	(0.347)
Observations	31671	
F	22.74	
df_m	167	
df_r	31670	

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Frequencies of One-time Movers by the Age at Move

Age at move	Count
2	556
3	919
4	1131
5	1287
6	1344
7	1370
8	1186
9	1212
10	1180
11	1211
12	1165
13	1242
14	1099
15	932
16	1950
17	1860
18	3337
19	20733
20	31360
21	23607
22	15099
23	12725
24	11925
25	10190
26	8151
27	6119
28	4658
29	3460
30	2516
31	1849
32	1346
33	1047
34	735
35	523
36	316
37	164
Total	179504

Table 15: Regression 7 Results

	All	Low Income	Boys	Low Income Boys	Girls	Low Income Girls
Exp. Effect	0.00240	0.0212	-0.00600	-0.0112	0.0106	0.0474*
	(0.792)	(0.172)	(0.632)	(0.623)	(0.390)	(0.023)
Obs.	368130	127044	80390	25280	93102	31671
F	175.2	70.97	49.86	19.45	50.64	25.72
df_m	346	143	346	143	346	143
df_r	368129	127043	80389	25279	93101	31670

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Frequencies of Permanent Residents by Region and Income Decile

	1	2	3	4	5	6	7	8	9	10	Total
Etelä-Karjala	1167	1424	1651	1779	1697	1594	1487	1115	702	397	13013
Etelä-Pohjanmaa	2719	3589	3832	3443	3121	2432	1775	1185	768	541	23405
Etelä-Savo	1599	1918	1972	1898	1677	1423	1028	776	520	338	13149
Kainuu	936	1378	1330	1188	945	730	519	370	209	126	7731
Kanta-Häme	1390	1904	2013	2017	2129	1949	1726	1302	836	456	15722
Keski-Pohjanmaa	873	1183	1103	1164	974	726	607	389	225	118	7362
Keski-Suomi	3167	4048	4307	4040	3541	3230	2677	2064	1429	872	29375
Kymenlaakso	1618	2051	2264	2284	2314	2338	2069	1610	1102	518	18168
Lappi	2368	3066	3109	2724	2595	2114	1610	1182	706	346	19820
Pirkanmaa	4016	5525	6228	6263	6398	5938	5328	4389	3262	2302	49649
Pohjanmaa	1505	2651	3070	3160	3124	2862	2350	1926	1273	722	22643
Pohjois-Karjala	2786	2966	3029	2658	2252	1815	1314	975	577	370	18742
Pohjois-Pohjanmaa	5451	8229	8245	7285	6379	5524	4293	3157	2019	1406	51988
Pohjois-Savo	3276	3824	3805	3629	3175	2820	2176	1600	1043	785	26133
Päijät-Häme	2300	2789	2718	2726	2347	2133	1733	1229	818	577	19370
Satakunta	3209	3417	3493	3574	3193	3042	2561	1987	1182	677	26335
Uusimaa	9137	13340	17587	18883	19222	19975	20661	20731	20938	23251	183725
Varsinais-Suomi	4074	5481	6245	6754	6808	6808	6219	5033	3710	2922	54054
Total	51591	68783	76001	75469	71891	67453	60133	51020	41319	36724	600384
Observations	600384										

Table 17: High School Matriculation Rates of Permanent Residents by Region and Income Decile

	1	2	3	4	5	6	7	8	9	10
Etelä-Karjala	.1919	.1896	.2296	.2422	.2575	.3092	.3584	.3794	.4316	.5063
Etelä-Pohjanmaa	.1982	.2098	.2396	.2785	.3150	.3586	.4175	.4253	.4388	.55823
Etelä-Savo	.2145	.1997	.2465	.2998	.3023	.3486	.3521	.3930	.45	.4556
Kainuu	.1560	.1850	.2113	.2668	.2952	.3041	.3719	.3784	.4450	.5079
Kanta-Häme	.1892	.1896	.2062	.2523	.2752	.3022	.3441	.3871	.4629	.5110
Keski-Pohjanmaa	.2176	.2375	.2602	.2947	.3306	.3733	.4036	.4036	.4533	.4068
Keski-Suomi	.2356	.2307	.2656	.2953	.3496	.3796	.4236	.4763	.5241	.5722
Kymenlaakso	.1656	.1697	.2142	.2167	.2610	.2836	.2953	.3137	.3675	.4459
Lappi	.2247	.2185	.2454	.2823	.3356	.3623	.3938	.4239	.4363	.5838
Pirkanmaa	.2774	.2695	.3036	.3425	.3642	.4124	.4574	.5218	.6024	.6755
Pohjanmaa	.2512	.2350	.2710	.3418	.3630	.4008	.4340	.4621	.5106	.6094
Pohjois-Karjala	.2344	.2576	.2991	.3345	.3819	.4259	.4939	.5190	.5477	.5108
Pohjois-Pohjanmaa	.2236	.2454	.2820	.3304	.3872	.4379	.4934	.5622	.6196	.6842
Pohjois-Savo	.2146	.2089	.2602	.3026	.3361	.3550	.4136	.45	.4851	.5312
Päijät-Häme	.1926	.1911	.2358	.2667	.3038	.3282	.3791	.4223	.4474	.5217
Satakunta	.2026	.2291	.2545	.2854	.3169	.3366	.3979	.4701	.4992	.5273
Uusimaa	.3678	.3453	.3574	.3972	.4494	.4886	.5411	.6057	.6896	.7872
Varsinais-Suomi	.2965	.2936	.3159	.3511	.3884	.4254	.4851	.5367	.6175	.7071
Total	.2516	.2541	.2880	.3289	.3713	.4116	.4664	.5275	.6121	.7223
Observations	51591	68783	76001	75469	71891	67453	60133	51020	41319	36724

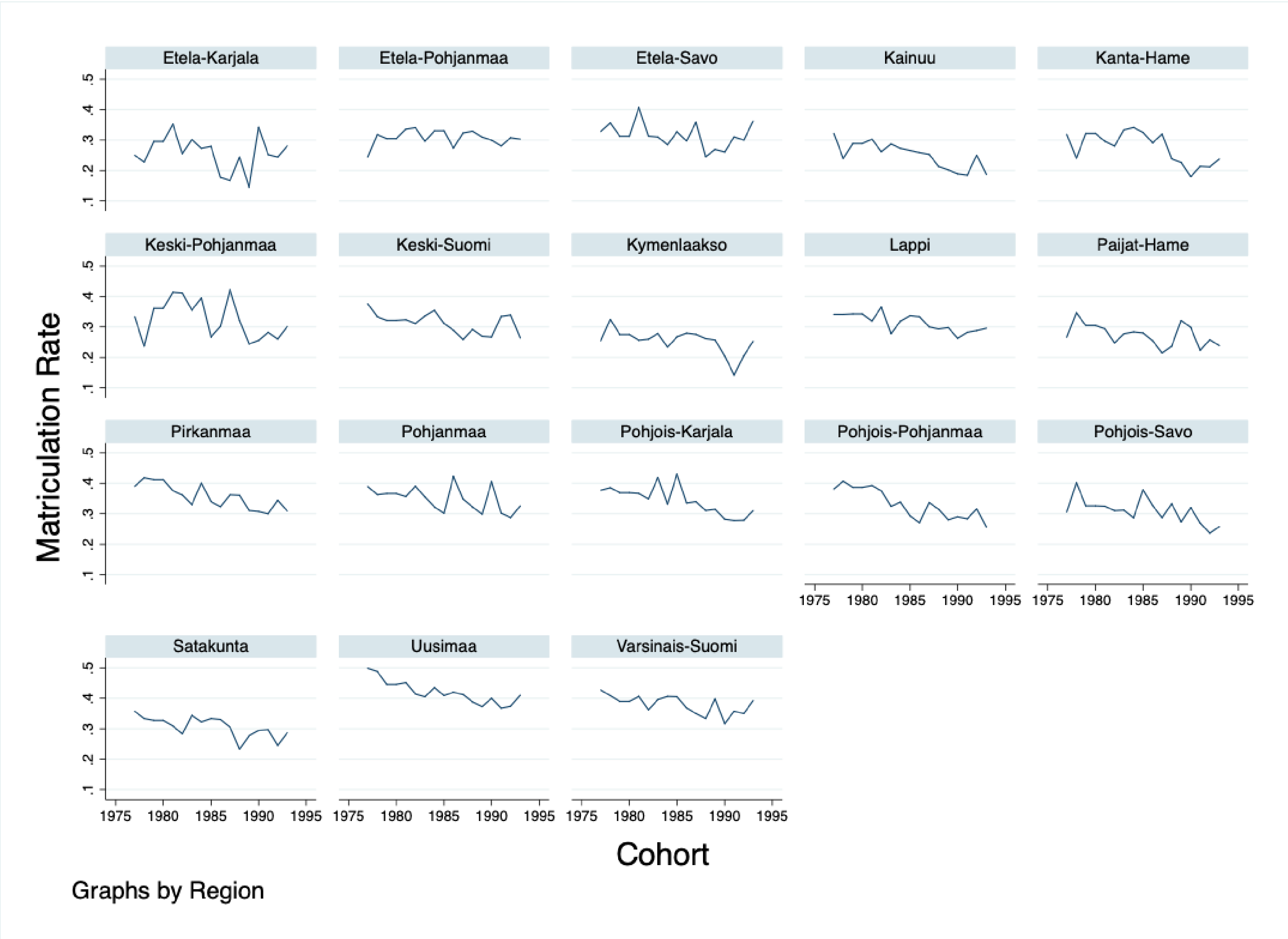

Figure 9: Matriculation Rates by Birth Cohort and Region